Analysis, Requirements and Development of a Collaborative Social and Medical Services Data Model

Risa B. Bobroff M.S., Cynthia A. Petermann M.S.,
J. Robert Beck M.D., and Gregory J. Buffone Ph.D.

Medical Informatics and Computing Research Program
Baylor College of Medicine
Houston, Texas

ABSTRACT

In any medical and social service setting, patient data must be readily shared among multiple providers for delivery of expeditious, quality care. This paper describes the development and implementation of a generalized social and medical services data model for an ambulatory population. The model, part of the Collaborative Social and Medical Services System Project, is based on the data needs of the Baylor College of Medicine Teen Health Clinics and follows the guidelines of the ANSI HISPP/MSDS JWG for a Common Data Model. Design details were determined by informal staff interviews, operational observations, and examination of clinic guidelines and forms. The social and medical services data model is implemented using object-oriented data modeling techniques and will be implemented in C++ using an Object-Oriented Database Management System.

INTRODUCTION

Delivery of social and medical services should be community based, providing ready access to the segment of the population served. Attainment of this goal requires an organizational infrastructure and computing and communication system capable of supporting geographically distributed clinics. To support these needs and to anticipate future expansion of distributed ambulatory services, the data model developed for the Baylor College of Medicine Teen Health Clinics (THCs) must be easily extensible and as well as serve the immediate needs of the clinics.

The Teen Clinics are representative of a medical and social care clinic. Spanning five sites throughout the city of Houston, Texas, the clinics provide a variety of free services to teenagers in Harris County. The services include family planning, Sexually Transmitted Disease (STD) screening and treatment, Case Management, patient education and counseling, school physicals, and prenatal and postpartum care. The clinics exhibit many of the problems typically associated with medical and/or social service clinics: patients may visit any of the clinic sites, making continuity of care difficult; clinic specific and aggregate statistical data are time consuming to acquire; manual and electronic data interchange with external entities are complex and time consuming; redundant and loosely coordinated data entry requires manual reconciliation of inconsistencies and is prone to errors and omissions.

The data modelling and implementation described here are integral parts of the Collaborative Social and Medical Services System (CSMSS) Project. The CSMSS is intended to provide the computing and communications support to fully integrate the Baylor College of Medicine Teen Health Clinics and address the problems enumerated above. We selected the THCs as the initial development domain for this data model due to the established needs of the clinics and the feasibility of designing an appropriate model. The THC environment is large enough to be representative of clinical and social service delivery, yet small enough to be understood and addressed in a reasonable time frame.

This paper describes a bottom-up process for developing a generic outpatient medical and social services data model. We present an evaluation of the high level data model and a strategy for testing the model in the clinics.

BACKGROUND AND RELATED WORK

Database Systems

Earlier systems, including the Clinical Information System (CIS) at Columbia Presbyterian Medical Center and the Virtual Database System (VDB) at Harvard Medical School Children's Hospital, used a relational model for their database implementation [1, 2, 3]. However, these system developers encountered a number of problems related to the difficulty of using relational tables to represent clinical data, including: negative performance impact from the use of full normalization; the need to keep table definitions constant over time despite chang-

0195-4210/94/$5.00 © 1994 AMIA, Inc.

ing data needs; and the challenge of displaying a meaningful view of the data to the user. In particular, the CIS project developers devised a generic extension table for each standard table in the database in order to accommodate differences in clinical findings. The VDB project required developers to denormalize data and store more than one type of data in the same table to achieve performance goals. They also mapped their hierarchical data representation to a set of tables that could be joined together, of necessity hiding the intuitive hierarchical nature of the data.

Object-Oriented Database Management Systems (OODBMS) are inherently well suited for representing and manipulating complex data models since the database stores data as persistent objects using the programming language's object representation. This storage mechanism alleviates the need for the compromises discussed above. OODBMSs offer more flexible mechanisms for schema evolution, for class inheritance, and for the introduction of new, related data types. These features are important in a generalized medical and social model as they make it easier for institutions to add new information to the data model, such as new tests or procedures. Furthermore, OODBMSs have the advantage of embedding data relationships within the model using direct links, as opposed to progammatical or dynamic relationships via relational joins, hence improving performance and the ability to present a logical view of the data model [4]. This implementation detail expedites many database operations, such as reviewing a list of recent test results for a patient. Rather than searching through the entire table of test results for a particular patient, there is a direct link from a patient to the corresponding test results.

JWG

The American National Standards Institute (ANSI) Health Informatics Standards Planning Panel (HISPP) / Message Standards Developers Subcommittee (MSDS) Joint Working Group (JWG) for a Common Data Model (CDM) is working to define and maintain a common data model for the health care industry. The JWG has published a Framework document that describes the notation and representation of the CDM and has generated a High-Level Data Model (HLDM). The HLDM consists of two diagrams: Subject Areas, showing the grouping of high level objects by subject; and Major Relationships, illustrating the relationships between the high level objects. This high level model was developed abstractly, i.e. topdown, by the JWG and is intended to serve as a guideline for model developers. A major goal of the JWG is to build a repository of the combined data models of each of the Standards Developing Organizations (SDOs), such as HL-7 and ACR-NEMA, in an effort to unify the data models of the independent SDOs and ease inter-standard communication [5].

METHODOLOGY

We used a number of sources to determine the elements in the data model. These included observations of operations and informal interviews with clinic staff members; the clinic Medical Protocols documents; the set of forms filled out by the staff; the reports generated by the clinics; discussion groups with representative clinic staff members; and the JWG HLDM subject definitions [5].

The informal interviews, observations of staff members, and examination of forms focused on data needs, storage (in the patient chart, in a notebook, etc.), and access. This input was used to define the classes, attributes, and relationships between classes in the model. For example, a number of objects were defined from the contents of the Patient Registration form including *PersonalData*, *EducationHistory*, *EmploymentHistory*, and *EmergencyContact*. The Medical Protocols helped enumerate the services provided by the clinics, define the attributes of each service class, determine default values and data ranges for those attributes, and describe typical services performed based on purpose of visit. A typical case is a six month visit during which the health care provider screens the patient for oral contraceptive side-effects, collects vital signs, collects specimens for gonorrhea, chlamydia and syphilis tests, discusses safe sex with the patient and reviews the proper use of the contraceptive method being used.

We used the JWG HLDM subject definitions as an initial guideline for our model development. We did not use the objects or the relationships between objects specified in the HLDM Major Relationships diagram in our development process, as we wanted to derive our own representation based on the Teen Clinics. Consequently, our model does not include all of the subject areas and high level classes specified in the JWG HLDM, as we, like the SDOs, are only developing the parts of the model that are within our application domain. We used the JWG recommended notation, a modified version of Coad and Yourdon [6], for representing our data model.

Several design goals strongly influenced the model's development. First, access control to the system is based on user role, with exceptions. To keep the access granularity high, classes are designed so that every attribute within a class may be viewed by all members of designated roles. For example, the *PersonalData* class contains a person's name, social security number, address(es) and phone number(s). A researcher gathering statistical data would not have access to this data. The write-once semantics of the database also influenced the design. In practice, the clinics amend data by crossing out the old data and replacing it with new data. Correspondingly, each database object instance is written once, and, aside from dynamically adding links to other instances, will never be updated. The data model will support amending data. Another design goal is utilizing semantic meaning in the data. For example, different test results appear in different formats. Rather than store all of the results as a generic string, each test is represented as a separate class, all of which are subclasses of the class *TestService*. Since the model is being implemented in C++, each test subclass can provide a method that translates its results to a common format for generalized handling. By keeping each result type separate, the user will be able to perform result specific comparisons on the data. Finally, the application data model contains attribute data types, while a separate data structure stores default value and range specifications for data in the application data model [5]. Separating these two types of data allows for greater domain independence, as each user may specify the dictionary(ies) to be used at any time. In addition, one clinic may easily maintain values and defaults different from the other clinics. This feature is important to the Teen Clinics, as the default populations the clinics serve differ by clinic location.

IMPLEMENTATION

The data model is being implemented in C++ using an OODBMS. We believe C++ and the OODBMS will ultimately facilitate extension of the model within the current domain or to other domains, provided the new domains follow the generalized structure defined for JWG HLDM and the THC. For example, during the model development process, the clinics added a number of new questions to their patient history form regarding sexual, physical, and mental abuse. To support this new function, an *AbuseHistory* class was added to the data model. When adding a new domain, the corresponding new tests and patient history classes may be added as subclasses or components of existing classes. Each new class inherits or implements a default data access interface from its superclass. In particular, if a predefined query listed all pending test results for a patient, adding a new test subclass would automatically add these tests to the search context, without requiring any changes to system code.

Another feature provided by the OODBMS is simplified data migration. In the case of the Teen Clinics, patients are not seen for general care after their 20th birthday. The OODBMS provides a mechanism by which we can migrate data on older patients to a different database. Longitudinal studies of clinic data may still access the relocated data by explicitly connecting to the archive database.

One significant drawback to OODBMSs is that object-oriented query languages are not well developed. Although some vendors provide an SQL interface, it is usually quite limited and data model navigation and queries must be implemented in the development language. This limitation makes user adhoc queries difficult to handle. However, a common object database standard has been proposed and published, and is supported by the major OODBMS vendors [7]. In the future, this standard may provide a common query language for OODBMSs. Despite this potential, a drawback to both types of systems remains -- the user must understand the entire data model to be able to query it. However, unlike relational systems, in which the attributes used to join tables would have to be explicitly stated externally to the database system, OODBMSs embed the relationships between classes in the model itself. This characteristic can also be a drawback to OODBMSs as the embedded relationships impose a structure on the data elements that is not present in relational systems.

EVALUATION

Comparison with the JWG HLDM

We generated a high level view of our data model and compared it to the data model presented by the JWG. Figure 1 shows how the two models compare. Since we divided our classes into subject areas based on the JWG HLDM, we expected our model to be a subset of the JWG model. However, until this figure was examined, we could not fully anticipate how the relationships we defined would compare with the JWG. We generated our high level model for comparison by extracting all of the classes contained in our model that are included on the JWG HLDM. We then compared the relationships between those classes contained in each model. We made one abstraction: since the Teen Clinics work minimally with finances and claims, we abstracted all of our financial classes into

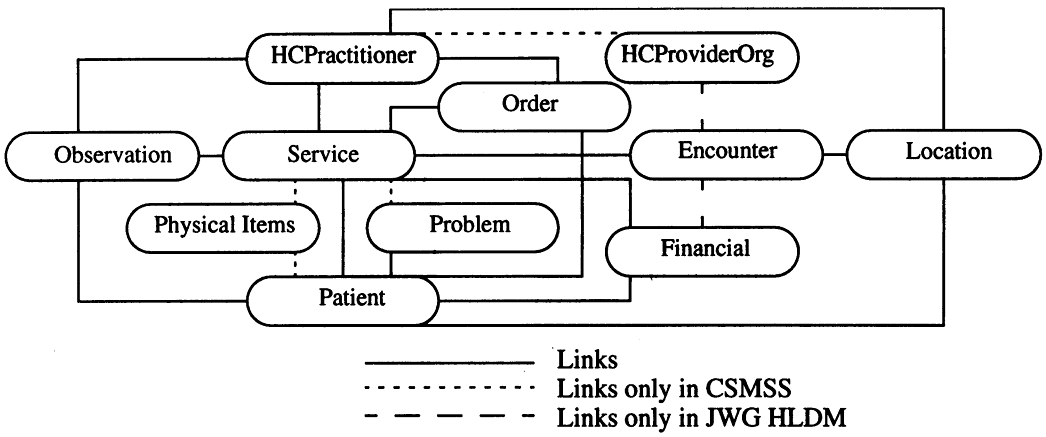

Figure 1. High level data model.

one representative *Financial* class, and did the same for the JWG HLDM.

The models are quite similar, as one would expect, given the thoughtful preparation of both models. This finding is also reassuring as it confirms that generalized models will be useful for system design and implementation. With a few exceptions, all of the connections between high level classes matched. One exception is the JWG HLDM relationship between *Encounter* and *Financial.* This relationship did not exist in our model because the claim forms used by the Teen Clinics are based on services, not encounters. Since encounter information can be accessed by traversing the relationship from service back to encounter, any necessary encounter information is still accessible using our model. Similarly, the JWG HLDM contains this same relationship, between *Service* and *Encounter*, and *can* therefore also derive encounter information from services.

The other major difference between the models is our relationship between *Problem* and *Service.* One might interpret this relationship as implicit in the JWG HLDM for the following reason. In the JWG HLDM subject definitions, *Problem* is contained in the *Observations* subject, and an *Observation* is reported to be the result of one or more *Services*. However, this relationship is not explicit in the JWG HLDM. Our research implied that an assessment service resulted in a problem definition, resulting in our adding a direct relationship between the two classes.

Our model contains one relationship not present in the JWG HLDM, that is, the relationship between *Health Care Practitioners* and *Health Care Provider Organization.* In our model, a health care practitioner belongs to one or more health care provider organizations. Also, our model includes some *relationships* not yet specified in the JWG HLDM: *Physical Items* to *Service* and *Physical Items* to *Patient.*

Comparison with Data Collected

Before review with clinic staff members, we inspected every form and report used by the clinics to ensure that every data item is represented in or can be derived from the data model. We also verified that every data item in the data model was used by a form or report. These checks were performed by generating a list of data items, sorted by source. The contents of this list was then compared to the data model, ensuring complete coverage. Then, each attribute of the data model was checked for use in the textual list.

Comparison with THC Operations

The model was reviewed with the clinic staff by walking through each form for each type of clinic visit, ensuring that we knew about all the patient data collected on a form (in the margins, etc.). We also distributed to each staff member a list of tasks performed, forms used, and report contents based on job description. Staff members returned the lists with their feedback.

Our intention is to deploy CSMSS initially at two sites, evaluate its performance, and then install the system at the remaining sites. During the phased deployment the clinics will maintain the small database system currently in use. At the end of each month, the monthly report will be generated using both systems and compared.

Before installing CSMSS in the clinics paper based versions of the forms will be used to collect clinic data. After clinic hours, staff members will enter the data from these forms into CSMSS. After data entry is completed, the standard clinic practices will be performed on this data: a nurse will audit the charts; a staff member will enter lab data; the billing clerk will generate claim forms from the new data; and a clerk will produce a monthly report. This investigation will be conducted to ensure that the format and content of the data stored by CSMSS fulfills the Clinics needs.

CSMSS is intended to be an active test environment with subsequent development of end user tools and multimedia facilities. Furthermore, feedback from the staff on the entire system and the data being stored by the system will be used to continuously improve the system.

Comparison with Other Domains
Other clinics affiliated with Baylor College of Medicine have expressed interest in the data model. After the Teen Clinic implementation is completed, we plan to evaluate the generality of the model by extending it to the domains of these additional clinics.

CONCLUSION

A number of data modelling questions arose during the design. These included: determining when to use an attribute to differentiate similar objects versus when to subtype the class; deciding when to create a direct relationship between two classes and when to use transitivity; deciding when to use an attribute to explicitly state a value, and when to use a link to an existing class (e.g. specify the name of the staff member who performed a service, or link to that staff member object); specifying mutual exclusivity in the data model using links (e.g. a patient may have a genitourinary exam or a pelvic exam, but not both). Comparison to analogous models which have been tested by implementation and use would be helpful in resolving these issues.

It is beyond the scope of this model to fully incorporate laboratory or other ancillary support models. Nevertheless, it would be advantageous to have generalized models for these areas that could be referenced or linked for purposes of completeness and implementation. The development of such data models and the means for referencing them for design and implementation purposes are needed. For example, a detailed model of the laboratory process and data could be combined with this model, thereby eliminating redundant terminology and conflicts in representation that arise from independent representations of the data relationships. We expect that our model will evolve concurrently with the CDM.

We are exploring methods of presenting the data model to the user to allow for ad-hoc queries and to eventually provide the user the means to extend the data model, as necessary.

In this paper we present the process followed for bottom-up development and implementation of an extensible data model for an outpatient medical and social service environment. We compared our model with the HLDM from the JWG. We are in the process of implementing the model and expect the use of an OODMS to accelerate implementation and improve maintenance efficiency and extensibility as we apply our system architecture and data model to other clinical domains.

Acknowledgments
The work reported here was supported in part by grant LM04905 from the National Library of Medicine. We thank the staff of the Baylor College of Medicine Teen Health Clinics for their help in this project.

References

[1]. C. Friedman, G. Hripcsak, S. B. Johnson, J. J. Cimino P. D. Clayton. A Generalized Relational Schema for an Integrated Clinical Patient Database. Proc. of the 14th Annual SCAMC, Wash., D.C., 1990.

[2]. S. Johnson, C. Friedman, J. J. Cimino, T. Clark, G. Hripcsak, P. Clayton. Conceptual Data Model for a Central Patient Database. Proc. of the 15th Annual SCAMC, Wash., D.C. 1991.

[3]. R. W. Stahlhut, D. P. McCallie, Jr., D. M. Waterman, D. M. Margulies. A Relational Model for Clinical Objective Results. Proc. of the 15th Annual SCAMC, Wash., D. C., 1990.

[4]. R. G. G. Cattell, Object Data Management: Object-Oriented and Extended Relational Database Systems. Addison-Wesley Publishing Company, Reading, MA, 1991.

[5]. ANSI/HISPP MSDS JWG for a Common Data Model IEEE P1157 Medical Data Interchange Working Group. Trial-Use Standard for Healthcare Data Interchange -- Information Model Methods. June, 1994.

[6]. P. Coad, E. Yourdon. Object Oriented Analysis, 2nd Edition. Yourdon Press, Englewood Cliffs, NJ, 1991.

[7]. R. G. G. Cattell Editor, The Object Database Standard: ODMG-93. Morgan Kaufmann Publishers, San Mateo, CA, 1993.